Supporting Information

Hsia et al. 10.1073/pnas.1006297107

SI Text

SI Methods. ***Protein expression and purification.*** A DNA fragment of yeast α-COP, encompassing residues 900–1,201, was amplified by PCR and cloned into a modified pET28a vector (1) that contained a PreScission protease site directly after the N-terminal hexa-histidine tag using the NdeI and NotI restriction sites. The resulting α-COP fusion proteins contained an N-terminal hexa-histidine tag with sequence MGSSHHHHHHSSGLE-VLFQGPH. The amplified PCR fragment of full-length yeast ϵ-COPI (residues 1–296) was cloned into the pET-Duet1 vector (Novagen) using BamHI and NotI restriction sites. A DNA fragment encoding yeast Dsl1, residues 410–440, was cloned into pGEX-TT (GE Heathcare) using BamHI and HindII restriction sites and was a kind gift from Hans Schmitt (2).

For the expression of ϵ-COP and the α-COP^{CTD}•ϵ-COP complex, *E. coli* BL21-CodonPlus (DE3)-RIL (Stratagene) were cotransformed with the appropriate expression vectors. Protein expression was carried out in LB media and induced by the addition of 0.5 mM isopropyl-β-D-thio-galactoside (IPTG) at 18 °C for 16 h. Cells were pelleted by centrifugation and resuspended in a buffer containing 50 mM K-phosphate, pH 7.4, 150 mM NaCl, 5 mM β-mercaptoethanol (β-ME), and Complete EDTA-free protease inhibitor cocktail tablets (Roche). The cells were lysed with a cell disrupter (Avestin), and the lysate was centrifuged for 90 min at 40,000 x *g*. The lysate was then applied onto a HIS-Select column (Sigma) and eluted via an imidazole gradient. Protein-containing fractions were pooled and dialyzed using a buffer containing 20 mM Hepes, pH 7.4, 50 mM NaCl, and 3 mM DTT. The protein was then applied to a HiTrap Q column (GE Healthcare) that was equilibrated in the same buffer and eluted with a NaCl gradient (50–1,000 mM). Peak fractions were concentrated and loaded onto a Superdex 200 10/300 GL gel filtration column (GE Healthcare) that was equilibrated with a buffer containing 20 mM Hepes, pH 7.5, 50 mM NaCl, and 3 mM DTT. Protein-containing fractions were then applied to a monoQ 5/50 GL column (GE Healthcare) and eluted via a NaCl gradient (50–500 mM) in a buffer containing 20 mM Hepes, pH 7.4, 50 mM NaCl, and 3 mM DTT. Fractions containing pure proteins were desalted using a HiTrap desalting column (GE Healthcare) that was equilibrated with a buffer containing 20 mM Hepes, pH 7.5, 50 mM NaCl, and 3 mM DTT. Purified ϵ-COP and the α-COP^{CTD}•ϵ-COP complex were concentrated to 30 mg/mL and 15 mg/mL, respectively and flash-frozen in liquid nitrogen for storage at −80 °C. The seleno-L-methionine (SeMet)-labeled α-COP^{CTD}•ϵ-COP complex was produced with a methionine inhibition pathway protocol (3).

The expression of the GST-Dsl1 peptide was carried out using the same protocol. However, the pelleted cells were resuspended in a buffer containing 20 mM Hepes, pH 7.5, 150 mM NaCl, 3 mM DTT, and Complete EDTA-free protease inhibitor cocktail tablets (Roche). For purification, the cells were lysed with a cell disruptor (Avestin), and the lysate was cleared by centrifugation for 90 min at 40,000 x *g*. The resulting supernatant fraction was filtered through a 0.45 μm filter (Nalgene) and loaded onto a Glutathione Sepharose 4B column (GE Healthcare) that was equilibrated in buffer A (50 mM Hepes, pH 7.5, 150 mM NaCl, and 3 mM DTT). The resin was first washed in buffer A, and the bound protein was eluted using a gradient of buffer B (buffer A supplemented with 20 mM glutathione). The purified GST-Dsl1 peptide was concentrated to 5 mg/mL and flash-frozen in liquid nitrogen for storage at −80 °C.

Crystallization and structure determination. Native crystals of the α-COP^{CTD}•ϵ-COP complex were obtained at 20 °C by vapor diffusion in hanging drops using 1 μL of the protein (6 mg/mL) and 1 μL of a reservoir solution consisting of 18% (*w*/*v*) PEG 3,350, 200 mM ammonium citrate, and 100 mM Hepes, pH 7.2. SeMet-labeled crystals were obtained at 20 °C by vapor diffusion in hanging drops using 1 μL of the protein (5 mg/mL) and 1 μL of a reservoir solution consisting of 20% (*w*/*v*) PEG 3,350, 200 mM NaCl, 100 mM Na-malonate, 8% (*w*/*v*) tacsimate, pH 6.0, and 100 mM Hepes, pH 7.2. Native and SeMet-labeled crystals appeared after three days and grew to their maximum size of 200 μm × 100 μm × 50 μm within two weeks.

For cryoprotection, native and Se-Met-labeled crystals were stabilized in 20% (*w*/*v*) PEG 3,350, 200 mM ammonium citrate, 100 mM Hepes, pH 7.2, and 23% (*v*/*v*) glycerol and 22% (*w*/*v*) PEG 3,350, 200 mM NaCl, 100 mM Na-malonate, 8% (*w*/*v*) tacsimate, pH 6.0, 100 mM Hepes, pH 7.2, and 22% (*v*/*v*) glycerol, respectively. Crystals were flash-frozen in liquid nitrogen-cooled liquid propane. X-ray diffraction data were collected at the General Medicine and Cancer Institutes Collaborative Access Team (GM/CA-CAT) beamline 23ID-B at the Advanced Photon Source (APS) at the Argonne National Laboratory (ANL). X-ray intensities were processed using HKL2000 (4), and the CCP4 program package (5) was used for subsequent calculations. A single-wavelength anomalous dispersion (SAD) X-ray diffraction dataset of a SeMet-labeled protein crystal was used to identify the positions of 24 selenium atoms with SHELXD (6). Phases were calculated to 3.25 Å in SHARP (7), followed by density modification in DM (5) with solvent flattening and histogram matching. This procedure yielded an electron density map of good quality. A model was built with the program O (8) and refined using CNS (9). No electron density was observed for the three C-terminal residues of ϵ-COP. These residues are presumed to be disordered and therefore have been omitted from the final model. The stereochemical quality of the model was assessed with PROCHECK (10) and MolProbity (11). There are no residues in the disallowed region of the Ramachandran plot. Data collection and refinement statistics are shown in Table S1.

Analytical ultracentrifugation. Sedimentation velocity experiments were performed at 4 °C in a Beckman Optima XL-I analytical ultracentrifuge. Double-sector cells were loaded with the protein sample (in a solution containing 50 mM Tris, pH 8.0, 150 mM NaCl, 2 mM Tris(2-carboxyethyl)phosphine (TCEP)) and the reference solution (50 mM Tris, pH 8.0, 150 mM NaCl, and 2 mM TCEP), respectively. Data were recorded with absorbance detection at a wavelength of 280 nm. The partial specific volume and the solvent density were calculated using the SEDNTERP program. The SEDFIT analysis program was used to analyze the absorbance profiles and to calculate the sedimentation coefficient distribution, c(s), which was then transformed into a molar mass distribution, c(M) (12).

Protein interaction analysis. Protein interaction experiments were carried out on a Superdex 200 10/300 GL gel filtration column (GE Healthcare) that was equilibrated in a buffer containing 20 mM Hepes, pH 7.5, 100 mM NaCl, and 3 mM DTT. Complexes were formed by incubating ≈1 mg of purified proteins for 1 h at 4 °C. Complex formation was monitored following the injection of the preincubated proteins as well as that of

the recombinant purified proteins in isolation. All proteins were analyzed under the same buffer conditions, and complex formation was confirmed by SDS-PAGE of the eluted fractions followed by Coomassie brilliant blue staining.

Illustration and figures. Figures were generated using PyMOL (13). The electrostatic potential was calculated using APBS (14). Sequence alignments were generated using ClustalX (15) and colored with Alscript (16).

1. Hoelz A, Nairn AC, Kuriyan J (2003) Crystal structure of a tetradecameric assembly of the association domain of Ca2+/calmodulin-dependent kinase II. *Mol Cell* 11:1241–1251.
2. Zink S, Wenzel D, Wurm CA, Schmitt HD (2009) A link between ER tethering and COP-I vesicle uncoating. *Dev Cell* 17:403–416.
3. Doublié S (1997) Preparation of selenomethionyl proteins for phase determination. *Methods Enzymol* 276:523–530.
4. Otwinowski Z, Minor W (1998) Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol* 276:307–326.
5. Collaborative Computational Project number 4 (1994) The CCP4 suite: programs for protein crystallography. *Acta Crystallogr D* 50:760–763.
6. Sheldrick GM (2008) A short history of SHELX. *Acta Crystallogr A* 64:112–122.
7. Bricogne G, Vonrhein C, Flensburg C, Schiltz M, Paciorek W (2003) Generation, representation and flow of phase information in structure determination: recent developments in and around SHARP 2.0. *Acta Crystallogr* 59:2023–2030.
8. Jones TA, Zou JY, Cowan SW, Kjeldgaard M (1991) Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr A* 47:110–119.
9. Brunger AT, et al. (1998) Crystallography & NMR System (CNS): a new software system for macromolecular structure determination. *Acta Crystallogr D* 54:905–921.
10. Laskowski RA, MacArthur MW, Moss DS, Thornton JM (1993) PROCHECK: a program to check the stereochemical quality of protein structures. *J Appl Crystallogr* 26:283–291.
11. Davis IW, et al. (2007) MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. *Nucleic Acids Res* 35:W375–383.
12. Schuck P (2000) Size-distribution analysis of macromolecules by sedimentation velocity ultracentrifugation and lamm equation modeling. *Biophys J* 78:1606–1619.
13. DeLano WL (2002) The PyMOL Molecular Graphics System (DeLano Scientific, San Carlos, CA).
14. Baker NA, Sept D, Joseph S, Holst MJ, McCammon JA (2001) Electrostatics of nanosystems: application to microtubules and the ribosome. *Proc Natl Acad Sci USA* 98:10037–10041.
15. Jeanmougin F, Thompson JD, Gouy M, Higgins DG, Gibson TJ (1998) Multiple sequence alignment with Clustal X. *Trends Biochem Sci* 23:403–405.
16. Barton GJ (1993) ALSCRIPT: a tool to format multiple sequence alignments. *Protein Eng* 6:37–40.

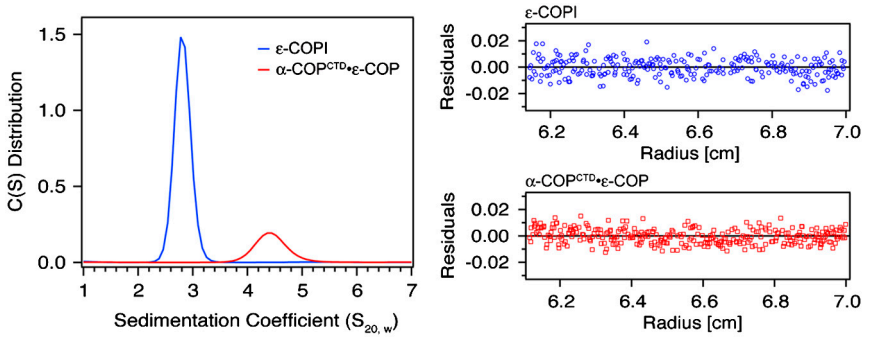

Fig. S1. Sedimentation velocity analyses of the α-COP^{CTD}•ε-COP complex and ε-COP. The sedimentation coefficient distribution shows single species for the α-COP^{CTD}•ε-COP complex and ε-COP (left). For the α-COP^{CTD}•ε-COP complex, the determined molecular mass is 69.2 ± 7.2 kDa, corresponding to a monomeric species of the α-COP^{CTD}•ε-COP heterodimer (calculated 71.9 kDa) in solution. For ε-COP, the determined molecular mass is 37.7 ± 3.1 kDa, corresponding to a monomeric species (calculated 35.4 kDa). No higher-order oligomers are detectable for either samples. The residuals for the fitting are shown on the right.

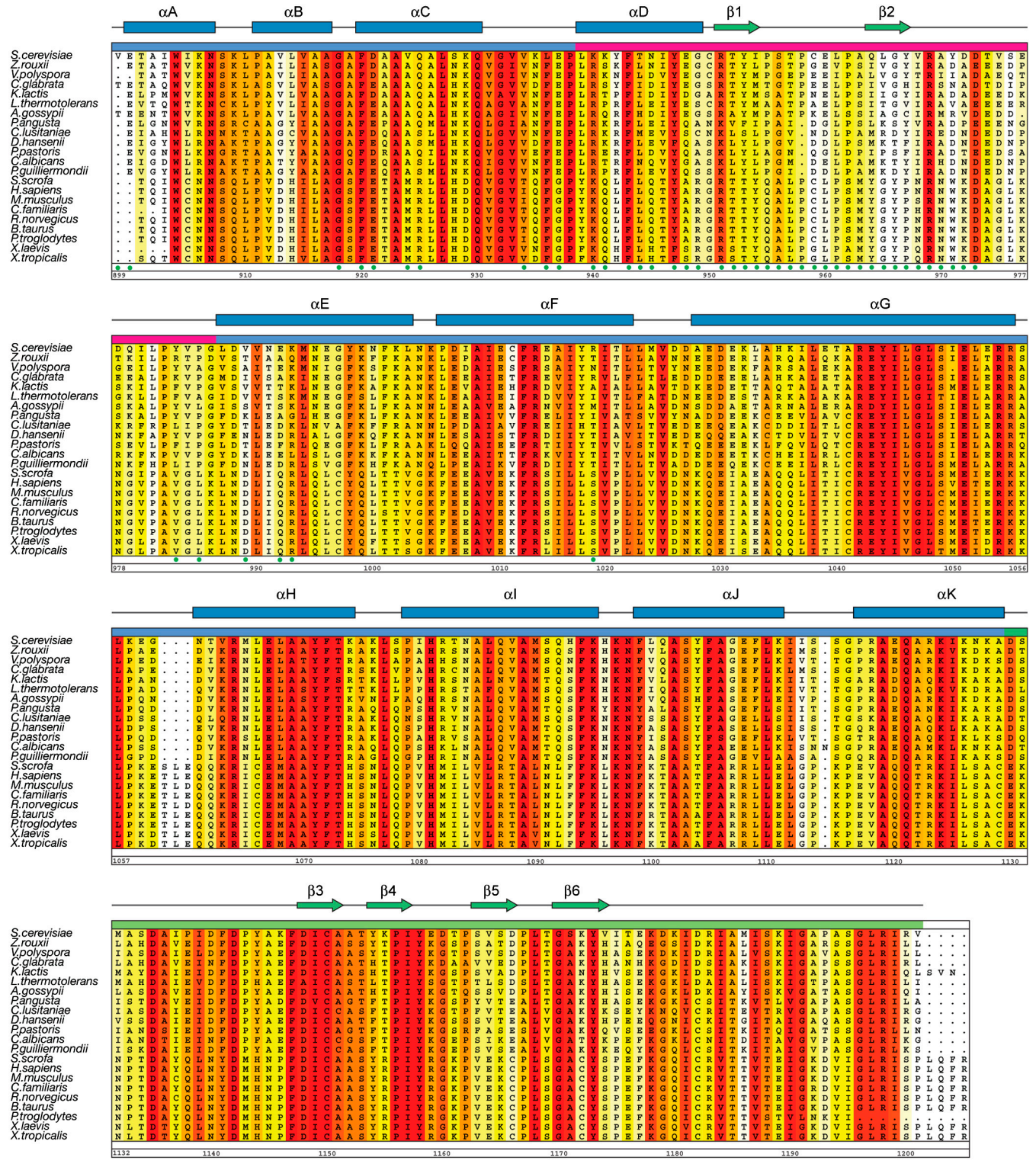

Fig. S2. Multispecies sequence alignment of α-COP C-terminal domain (CTD) homologs. The numbering of the residues and the secondary structure are according to yeast α-COP CTD. The secondary structure is indicated above the sequence as green arrows (β-strands), blue rectangles (α-helices), and gray lines (coil regions). The overall sequence conservation at each position is shaded in a color gradient from yellow (60% similarity) to red (100% identity) using the Blosum62 weighting algorithm (1). The α-COP CTD residues that are involved in the interaction with ε-COP are indicated with green dots below the alignment.

1 Henikoff S, Henikoff JG (1992) Amino acid substitution matrices from protein blocks. *Proc Natl Acad Sci USA* 89:10915–10919.

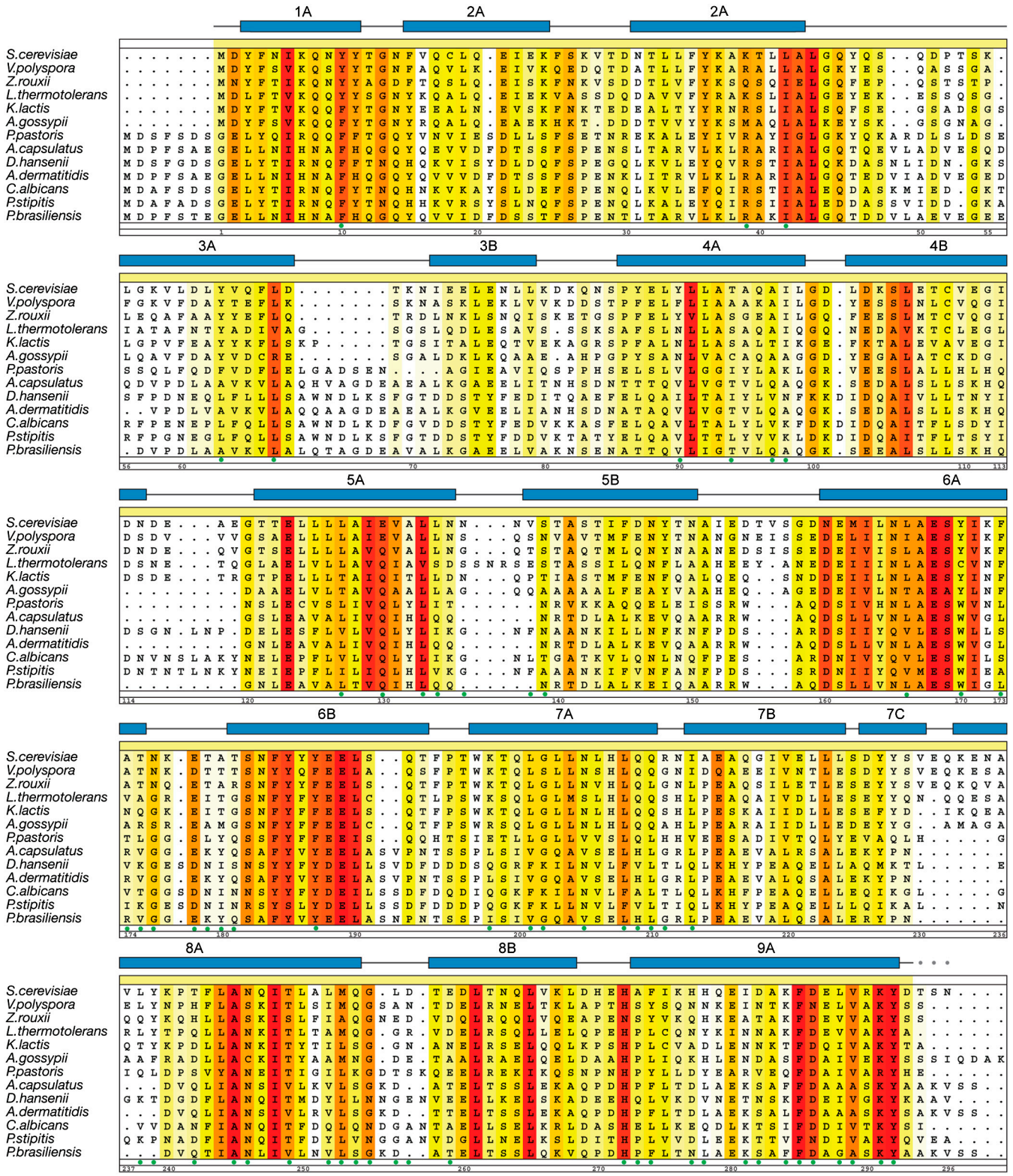

Fig. S3. Multispecies sequence alignment of ϵ-COP homologs. The numbering of the residues and the secondary structure are according to yeast ϵ-COP. The secondary structure is indicated above the sequence as blue rectangles (α-helices), gray lines (coil regions), and gray dots (disordered residues). The overall sequence conservation at each position is shaded in a color gradient from yellow (60% similarity) to red (100% identity) using the Blosum62 weighting algorithm (1). The ϵ-COP residues that are involved in the interaction with the α-COP CTD are indicated with green dots below the alignment.

1 Henikoff S, Henikoff JG (1992) Amino acid substitution matrices from protein blocks. *Proc Natl Acad Sci USA* 89:10915–10919.

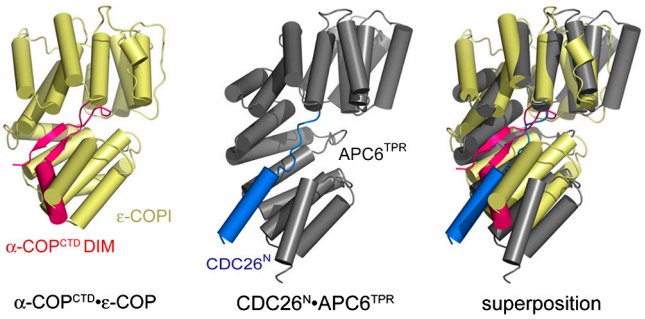

Fig. S4. Structural comparison between the α-COP^{DIM}•ε-COP and CDC26^{N}•APC6^{TPR} complexes. The α-COP^{DIM}•ε-COP (left) and the CDC26^{N}•APC6^{TPR} (middle) are illustrated in magenta and yellow and blue and gray, respectively. A superposition based on the TPRs of ε-COP and the APC6^{TPR} domain is shown on the right. The ε-COP TPRs substantially deviate from the canonical superhelical conformation observed in other TPR proteins (1).

1 Blatch GL, Laessle M (1999) The tetratricopeptide repeat: a structural motif mediating protein-protein interactions. *Bioessays* 21:932–939.

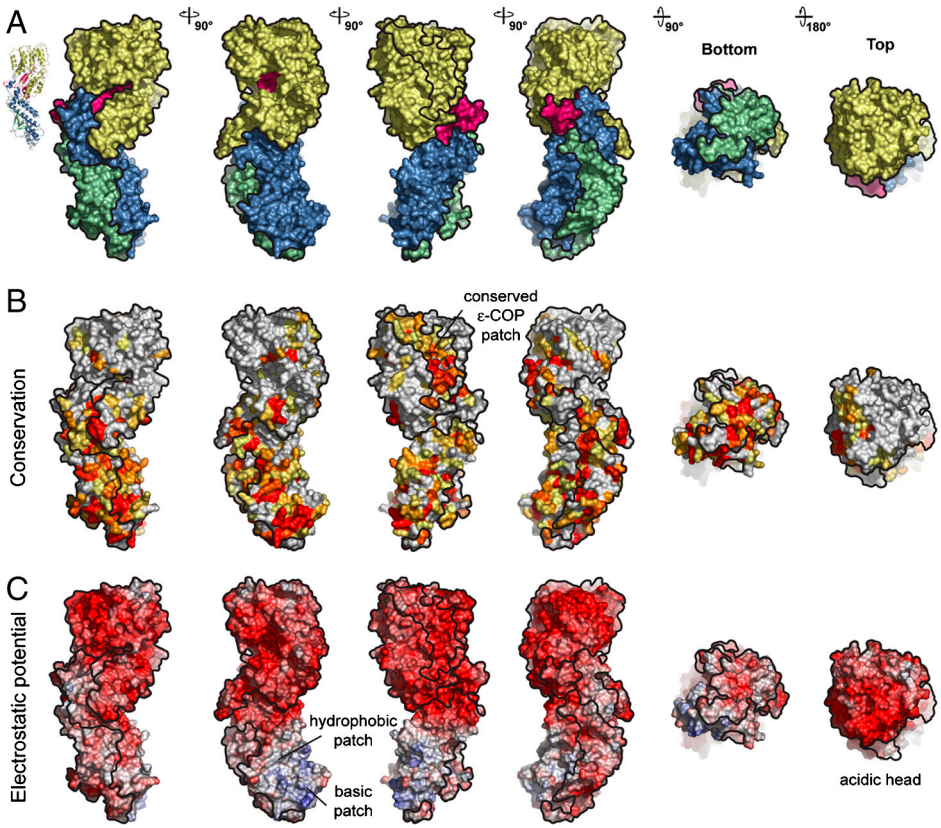

Fig. S5. Surface properties of α-COP^{CTD}•ε-COP. (*A*) Surface renditions of the α-COP^{CTD}•ε-COP heterodimer colored according to Fig. 1*B*. As a reference, a ribbon representation of the α-COP^{CTD}•ε-COP is shown and corresponds to the orientation on the left. (*B*) The surface representations of α-COP^{CTD}•ε-COP are colored according to multispecies sequence alignments (Figs. S2 and S3). The conservation at each position is mapped onto the surface and shaded in a color gradient from yellow (60% similarity) to red (100% identity). (*C*) α-COP^{CTD} • ε-COP surface renditions colored according to the electrostatic potential, ranging from red ($-10k_BT/e$) to blue ($+10k_BT/e$).

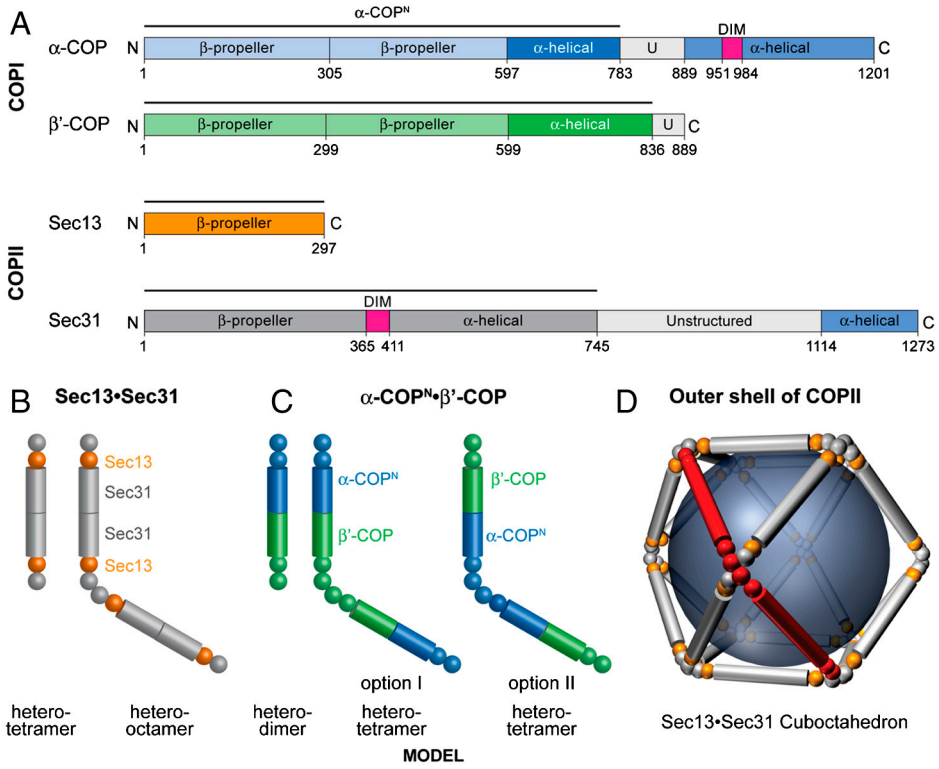

Fig. S6. Shared domain architecture of the outer layer COPI and COPII components. (*A*) Domain structures of yeast α-COP, β′-COP, Sec13, and Sec31. For α-COP, the two putative N-terminal β-propeller domains (light blue), the central α-helical domain (dark blue), the unstructured region (U, gray), and the C-terminal domain (steel blue) containing the domain invasion motif (DIM) (magenta) are indicated. β′-COP has the same domain structure as α-COP but lacks the C-terminal domain. The two putative N-terminal β-propeller domains and the central α-helical domain are colored in light and dark green, respectively. For Sec13, the six-bladed β-propeller is shown (orange). For Sec31, the N-terminal, seven-bladed, β-propeller domain (gray), the Sec13-binding DIM (magenta), the central α-helical domain (gray), the unstructured region (light gray), and the C-terminal domain that is dispensable for coat formation (steel blue) are indicated. The black lines above the domain structures indicate the parts of the proteins that form the structures in *B* and *C*. (*B*) Schematic representation of the Sec13•Sec31 heterotetramer and heterooctamer. The Sec31 DIM facilitates the association with Sec13. The central α-helical domain of Sec31 homodimerizes and forms a heterotetramer. Two heterotetramers form a heterooctamer, which is the repeating unit of COPII cages (1, 2). (*C*) Schematic representation of the architecture of a putative α-COP^{N}•β′-COP heterodimer and heterotetramer. The heterodimer is formed by the heterodimerization of the central α-helical domains of α-COP and β′-COP. Two heterodimers form a heterotetramer through the homodimerization of the N-terminal β-propeller domain of β′-COP or α-COP. β propellers and α-helical domains are represented by spheres and cylinders, respectively. (*D*) Sec13•Sec31 heterooctamers oligomerize into the outer coat of COPII with various geometries. Shown is the smallest cuboctahedral cage that is composed of 12 heterooctamers. The central α-helical domains form the edges of the COPII cage. Notably, the COPII vertices are formed by the asymmetric homotetramerization of the N-terminal β-propeller domain of Sec31. For COPI, there are two options for the formation of COPII-like vertices, depending on which N-terminal β-propeller (α-COP or β′-COP) is used in the formation of the heterotetramer. All other arrangements of α-COP and β′-COP would result in cages that are architecturally distinct from COPII.

1 Fath S, Mancias JD, Bi X, Goldberg J (2007) Structure and organization of coat proteins in the COPII cage. *Cell* 129:1325–1336.
2 Stagg SM, et al. (2008) Structural basis for cargo regulation of COPII coat assembly. *Cell* 134:474–484.

Movie S1.

Movie S1 (MOV)

Movie S2

Movie S2 (MOV)

Table S1. Crystallographic analysis

	Crystal 1 native	Crystal 2 SeMet
Data collection		
Protein	α-COP^{CTD}•ε-COP	α-COP^{CTD}•ε-COP
Synchrotron	APS*	APS*
Beamline	GM/CA CAT, 23ID-B	GM/CA CAT, 23ID-B
Space group	C2	C2
Cell dimensions		
a, *b*, *c* (Å)	$a = 329.1$, $b = 74.4$, $c = 97.2$	$a = 328.1$, $b = 74.3$, $c = 96.4$
α,β,γ (°)	$\alpha = 90.0$, $\beta = 102.3$, $\gamma = 90.0$	$\alpha = 90.0$, $\beta = 102.0$, $\gamma = 90.0$
		Se Peak
Wavelength (Å)	1.03316	0.9796
Resolution (Å)†	20.0-2.9 (3.0-2.9)	50.0-3.25 (3.37-3.25)
R_{sym} (%)†	10.2 (36.8)	10.5 (73.5)
$\langle I/\sigma I \rangle$†	10.5 (1.7)	10.3 (1.8)
Completeness (%)†	93.3 (65.6)	99.9 (99.9)
Redundancy	3.1	3.8
Refinement		
Resolution (Å)	20.0-2.90	50.0-3.25
No. reflections	51,235	62,887
R_{work}/R_{free} (%)	24.6/29.6	24.1/28.3
No. protein atoms	14,276	14,259
B-factors protein	86.4	85.2
R.m.s. deviations		
Bond lengths (Å)	0.011	0.010
Bond angles (°)	1.63	1.64
Ramachandran plot‡		
Most favored (%)	80.4	77.8
Additionally allowed (%)	16.9	19.2
Generously allowed (%)	2.8	3.0
Disallowed (%)	0.0	0.0

*APS, Advanced Photon Source, Argonne National Laboratory.
†Highest-resolution shell is shown in parentheses.
‡As determined by Procheck (10).